

CERTIFIED COPY OF
PRIORITY DOCUMENT



#4 37-91
Priority Papers
Jc825 U.S. PTO

09/725419



11/29/00

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 199 57 430.8

Anmeldetag: 30. November 1999

Anmelder/Inhaber: Philips Corporate Intellectual Property GmbH,
Hamburg/DE

Bezeichnung: Verfahren zur Erzeugung eines Maximum-Entropie-
Sprachmodells

IPC: G 10 L 15/00

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Anmeldung.

München, den 02. Juni 2000
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

12/10/00



PHD 99-177

BESCHREIBUNG

Verfahren zur Erzeugung eines Maximum-Entropie-Sprachmodells

Die Erfindung betrifft ein Verfahren zur Erzeugung eines Maximum-Entropie-Sprachmodells für ein Spracherkennungssystem.

- 5 Bei der Erzeugung von Sprachmodellen für Spracherkennungssysteme stellt sich das Problem, dass die Trainingskorpora nur begrenzte Mengen an Trainingsmaterial enthalten. Wahrscheinlichkeiten für sprachliche Äußerungen, die allein aus den entsprechenden Auftretshäufigkeiten im Trainingskorpora abgeleitet sind, werden deshalb Glättungsprozeduren
- 10 unterworfen, beispielsweise bei backing-off-Techniken. Doch nutzen backing-off-Sprachmodelle den statistischen Informationsgehalt von verfügbaren Trainingsdaten regelmäßig nicht optimal aus, da nicht gesehene Vergangenheiten von N-Grammen lediglich dadurch kompensiert werden, daß das jeweils betrachtete N-Gramm soweit verkürzt wird, bis man eine von Null verschiedene Auftretshäufigkeit im Trainingskorpora erhält. Mit Maximum-
- 15 Entropie-Sprachmodellen kann diesem Problem entgegengewirkt werden (vgl. R. Rosenfeld, "A maximum entropy approach to adaptive statistical language modeling", Computer, Speech and Language, 1996, Seiten 187-228). Mittels derartiger Sprachmodelle können insbesondere sowohl Auftretshäufigkeiten von N-Grammen als auch von Lücken-N-
- 20 Grammen im Trainingskorpora zur Schätzung von Sprachmodellwahrscheinlichkeiten verwendet werden, was bei backing-off-Sprachmodellen nicht der Fall ist. Allerdings tritt bei der Erzeugung eines Maximum-Entropie-Sprachmodells das Problem auf, geeignete Randwerte zu schätzen, von deren Wahl die iterierten Sprachmodellwerte des Maximum-
- 25 Entropie-Sprachmodells abhängen. Die Sprachmodellwahrscheinlichkeiten $p_{\lambda}(w|h)$ eines solchen Sprachmodells (w : Vokabularelement; h : Vergangenheits von Vokabularelementen bezüglich w) sind während eines Trainings so zu bestimmen, daß sie Randwertgleichungen der Form

$$m_{\alpha} = \sum_{(h,w)} p_{\lambda}(w|h) \cdot N(h) \cdot f_{\alpha}(h,w)$$

- möglichst gut erfüllen. Hierbei stellt m_α jeweils einen Randwert für eine a priori festzusetzende Bedingung α dar, von deren Erfüllung es abhängt, ob die Filterfunktion $f_\alpha(h, w)$ den Wert Eins oder den Wert Null annimmt. Eine Bedingung α fragt dabei ab, ob eine betrachtete Folge (h, w) von Vokabularelementen ein bestimmtes N-Gramm (der Begriff
- 5 N-Gramm soll auch Lücken-N-Gramme umfassen) ist oder auf ein bestimmtes N-Gramm ($N \geq 1$) endet, wobei N-Gramm-Elemente auch Klassen sein können, die in besonderem Bezug zueinander stehende Vokabularelemente zusammenfassen. $N(h)$ gibt die Auftretshäufigkeit der Vergangenheit h im Trainingskorpus an.
- 10 Unter allen Wahrscheinlichkeitsverteilungen, welche die Randwertgleichungen erfüllen, wird bei der Maximum-Entropie-Modellierung diejenige Verteilung ausgewählt, welche die bedingte Entropie

$$-\sum_h N(h) \sum_w p_\lambda(w|h) \log p_\lambda(w|h)$$

15

maximiert. Die spezielle Verteilung hat die Form

$$p_\lambda(w|h) = \frac{1}{Z_\lambda(h)} \exp \left\{ \sum_\alpha \lambda_\alpha f_\alpha(h, w) \right\} \quad \text{mit} \quad Z_\lambda(h) = \sum_{v \in V} \exp \left\{ \sum_\alpha \lambda_\alpha f_\alpha(h, v) \right\}$$

- 20 mit geeigneten Parametern λ_α .

- Bei der Iteration eines Maximum-Entropie-Sprachmodells wird insbesondere der sogenannte GIS-Algorithmus ("Generalized Iterative Scaling") eingesetzt, dessen Grundstruktur in J.N.Darroch, D. Ratcliff: "Generalized iterative scaling for log-linear models", The
- 25 Annals of Mathematical Statistics, 43(5), Seiten 1470-1480, 1972 beschrieben ist. Ein Ansatz zur Bestimmung der erwähnten Randwerte m_α beruht beispielsweise auf der Maximierung der Wahrscheinlichkeit des verwendeten Trainingskorpus, was auf Randwerte $m_\alpha = N(\alpha)$ führt, d.h. es wird ermittelt, wie häufig die Bedingungen α im Trainingskorpus erfüllt sind. Dies ist beispielsweise in S.A. Della Pietra, V. J. Della Pietra, J.

- Lafferty, "Inducing Features of random fields", Technical report, CMU-CS-95-144, 1995 beschrieben. Die Randwerte m_α erzwingen jedoch häufig, dass einige Sprachmodellwahrscheinlichkeitswerte $p_\lambda(w | h)$ der durch die Randwertgleichungen eingeschränkten Modelle verschwinden (d.h. zu Null werden), und zwar insbesondere für im Trainingskorpus nicht gesehene Folgen (h, w) . Verschwindende Sprachmodellwahrscheinlichkeitswerte $p_\lambda(w | h)$ sind aber aus zwei Gründen zu vermeiden: Einerseits könnte ein Spracherkennungssystem in solchen Fällen Sätze mit der Wortfolge (h, w) nicht erkennen, auch wenn diese plausible Erkennungsergebnisse wären, nur weil sie nicht im Trainingskorpus vorkommen. Andererseits widersprechen Werte $p_\lambda(w | h) = 0$ der funktionalen Gestalt der
- 10 Lösung aus der oben angeführten Gleichung für $p_\lambda(w | h)$, solange die Parameter λ_α auf endliche Werte beschränkt sind. Diese sogenannte Inkonsistenz (vgl. J.N.Darroch, D. Ratcliff am oben genannten Ort) verhindert die Lösung der Randwertgleichungen mit allen bisher bekannten Trainingsverfahren.
- 15 Der Erfindung liegt nun die Aufgabe zugrunde, ein Verfahren zur Erzeugung von Maximum-Entropie-Sprachmodellen so auszugestalten, das eine Verbesserung der statistischen Eigenschaften des erzeugten Sprachmodells erreicht wird.

Die Aufgabe wird dadurch gelöst, dass

- 20 durch Auswertung eines Trainingskorpus erste Wahrscheinlichkeitswerte $p_{ind}(w | h)$ für N-Gramme mit $N \geq 0$ gebildet werden;
- eine Schätzung von zweiten Wahrscheinlichkeitswerten $p_\lambda(w | h)$, die Sprachmodellwerte des Maximum-Entropie-Sprachmodells repräsentieren, in Abhängigkeit von den ersten Wahrscheinlichkeitswerten durchgeführt wird;
- 25 Randwerte m_α entsprechend der Gleichung

$$m_\alpha = \sum_{(h,w)} p_{ind}(w | h) \cdot N(h) \cdot f_\alpha(h, w)$$

- bestimmt werden, wobei $N(h)$ die Auftrittshäufigkeit der jeweiligen Vergangenheit h im Trainingskorpus und $f_\alpha(h, w)$ eine Filterfunktion ist, die nur für bestimmte durch den Index α gekennzeichnete und a priori vorgegebene N-Gramme einen von Null
- 30 verschiedenen Wert und ansonsten den Wert Null aufweist;

eine Iteration von Sprachmodellwerten des Maximum-Entropie-Sprachmodells solange durchgeführt wird, bis im n-ten Iterationsschritt nach der Formel

$$m_{\alpha}^{(n)} = \sum_{(h,w)} p_{\lambda}^{(n)}(w|h) \cdot N(h) \cdot f_{\alpha}(h,w)$$

ermittelte Werte $m_{\alpha}^{(n)}$ nach einem vorgebbaren Konvergenzkriterium hinreichend genau den Randwerten m_{α} angenähert sind.

5

Eine derartige Sprachmodellbildung führt auf ein Sprachmodell, welches die Statistik des Trainingskorpus besser auf die Statistik der zu erkennenden Sprache verallgemeinert, indem die Schätzung der Wahrscheinlichkeiten $p_{\lambda}(w|h)$ für ungesehene Wortübergänge (h, w) unterschiedliche Statistiken des Trainingskorpus ausnutzt: Neben den kürzerreich-

10

weitigen N-Grammen (wie bei backing-off-Sprachmodellen) können dann auch Lücken-N-Gramm-Statistiken und Korrelationen zwischen Wort-Klassen bei der Schätzung der Werte $p_{\lambda}(w|h)$ berücksichtigt werden.

15

Insbesondere ist vorgesehen, dass zur Iteration der Sprachmodellwerte des Maximum-Entropie-Sprachmodells, d.h. zum iterativen Training, der GIS-Algorithmus verwendet wird. Die ersten Wahrscheinlichkeitswerte $p_{\text{ind}}(w|h)$ sind vorzugsweise backing-off-Sprachmodellwahrscheinlichkeitswerte.

20

Die Erfindung bezieht sich auch auf ein Spracherkennungssystem mit einem entsprechend hergestellten Sprachmodell.

Ausführungsbeispiele der Erfindung werden im folgenden unter Verwendung einer Figur näher erläutert.

25

Die Figur zeigt ein Spracherkennungssystem 1, dem an seinem Eingang 2 in elektrischer Form vorliegende Sprachsignale zugeführt werden. Ein Funktionsblock 3 faßt eine akustische Analyse zusammen, die dazu führt, daß sukzessive am Ausgang 4 Merkmalsvektoren ausgegeben werden, die die Sprachsignale beschreiben. Bei der akustischen Analyse werden

30

die in elektrischer Form vorliegenden Sprachsignale abgetastet und quantisiert und anschließend in Rahmen zusammengefaßt. Aufeinanderfolgende Rahmen überlappen sich

- vorzugsweise dabei teilweise gegenseitig. Für jeden einzelnen Rahmen wird ein Merkmalsvektor gebildet. Der Funktionsblock 5 faßt die Suche nach der Folge von Sprachvokabularelementen zusammen, die bei der gegebenen Eingabe der Folge von Merkmalsvektoren am wahrscheinlichsten ist. Wie bei Spracherkennungssystemen üblich, wird hierzu die Wahrscheinlichkeit des Erkennungsergebnisses mit Hilfe der sogenannten Bayes-Formel maximiert. Hierbei geht sowohl ein akustisches Modell der Sprachsignale (Funktionsblock 6) als auch ein linguistisches Sprachmodell (Funktionsblock 7) in die Verarbeitung gemäß Funktionsblock 5 ein. Das akustische Modell gemäß Funktionsblock 6 beinhaltet die übliche Verwendung von sogenannten HMM-Modellen ("Hidden Markov Models") für die Modellierung einzelner Vokabularelemente oder auch einer Kombination mehrerer Vokabularelemente. Das Sprachmodell (Funktionsblock 7) enthält geschätzte Wahrscheinlichkeitswerte für Vokabularelemente oder Folgen von Vokabularelementen. Hierauf bezieht sich die im folgenden näher erläuterte Erfindung, die dazu führt, daß die Fehlerrate des am Ausgang 8 ausgegebenen Erkennungsergebnisses reduziert wird.
- 15 Außerdem wird die Perplexität des Systems reduziert.

- Beim erfindungsgemäßen Spracherkennungssystem 1 wird ein Sprachmodell mit Wahrscheinlichkeitswerten $p_{\lambda}(w | h)$, d.h. bedingten N-Gramm-Wahrscheinlichkeiten mit $N \geq 0$ für N-Gramme (h, w) verwendet (mit h als Vergangenheit aus $N-1$ Elementen bezüglich des Vokabularelementes w bezeichnet), das auf einem Maximum-Entropie-Ansatz beruht. Hierbei wird die gesuchte Verteilung durch bestimmte Randverteilungen eingeschränkt, und unter diesen Randbedingungen wird das Modell maximaler Entropie ausgewählt. Die Randbedingungen können sich sowohl auf N-Gramme verschiedener Länge ($N = 1, 2, 3, \dots$) als auch auf Lücken-N-Gramme beziehen, z. B. auf Lücken-
- 25 Bigramme der Form $(u, *, w)$, wobei $*$ ein Platzhalter für mindestens ein beliebiges N-Gramm-Element zwischen den Elementen u und w ist. Ebenso können N-Gramm-Elemente Klassen C sein, die Vokabularelemente zusammenfassen, die einen besonderen Bezug zueinander haben, z. B. indem sie grammatikalische oder semantische Verwandtschaften zeigen.

30

Die Wahrscheinlichkeiten $p_{\lambda}(w | h)$ werden in einem Training auf der Grundlage eines Trainingskorpus (z. B. NAB-Korpus - North American Business News) gemäß folgender

Formel geschätzt:

$$p_{\lambda}(w|h) = \frac{1}{Z_{\lambda}(h)} \exp \left\{ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}(h, w) \right\} \quad \text{mit} \quad Z_{\lambda}(h) = \sum_{v \in V} \exp \left\{ \sum_{\alpha} \lambda_{\alpha} f_{\alpha}(h, v) \right\} \quad (1)$$

- 5 Die Güte des so gebildeten Sprachmodells wird entscheidend von der Wahl von Randwerten m_{α} bestimmt, von denen die zu bestimmenden Wahrscheinlichkeitswerte $p_{\lambda}(w|h)$ für das Sprachmodell abhängen, was durch die folgende Formel ausgedrückt wird:

$$m_{\alpha} = \sum_{(h,w)} p_{\lambda}(w|h) \cdot N(h) \cdot f_{\alpha}(h, w) \quad (2)$$

10

Die Randwerte m_{α} werden mittels eines bereits berechneten und vorliegenden Sprachmodells mit den Sprachmodellwahrscheinlichkeiten $p_{\text{ind}}(w|h)$ geschätzt. Dazu wird Formel (2) verwendet, in der lediglich $p_{\lambda}(w|h)$ durch $p_{\text{ind}}(w|h)$ zu ersetzen ist, so daß eine Schätzung der m_{α} anhand der Formel

15

$$m_{\alpha} = \sum_{(h,w)} p_{\text{ind}}(w|h) \cdot N(h) \cdot f_{\alpha}(h, w) \quad (3)$$

erfolgt.

- 20 Die Werte $p_{\text{ind}}(w|h)$ sind insbesondere Wahrscheinlichkeitswerte eines auf der Basis des zugrundegelegten Trainingskorpus ermittelten sogenannten backing-off-Sprachmodells (siehe beispielsweise R. Kneser, H. Ney, "Improved backing-off for M-gram language modeling", ICASSP 1995, Seiten 181-185). Die Werte $p_{\text{ind}}(w|h)$ können aber auch anderen als gegeben vorauszusetzenden (schon berechneten) Sprachmodellen entnommen
- 25 werden, wie sie beispielsweise in A. Nadas: "Estimation of Probabilities in the Language Model of the IBM Speech Recognition System", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol. ASSP-32, Seiten 859-861, Aug. 1984 und in S.M. Katz: "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer", IEEE Trans. on Acoustics, Speech and Signal Proc., Vol. ASSP-35, Seiten 400-401,

March 1987, beschrieben sind.

$N(h)$ gibt die Häufigkeit der jeweiligen Historie h im Trainingskorpus an. $f_\alpha(h, w)$ ist eine zu einer Bedingung α korrespondierende Filterfunktion, die einen von Null verschiedenen

- 5 Wert (hier den Wert Eins) hat, wenn die Bedingung α erfüllt ist, und ansonsten gleich Null ist. Die Bedingungen α und die zugehörigen Filterfunktionen f_α werden heuristisch für das jeweilige Trainingskorpus festgelegt. Insbesondere wird hierbei ausgewählt, für welche Wort- oder Klassen-N-Gramme oder Lücken-N-Gramme die Randwerte fixiert werden.

10

Bedingungen α , bei denen $f_\alpha(h, w)$ den Wert Eins hat, sind vorzugsweise:

- ein betrachtete N-Gramm endet auf ein bestimmtes Vokabularelement w ;
- ein betrachtetes N-Gramm (h, w) endet auf ein Vokabularelement w , das einer bestimmten Klasse C angehört, die Vokabularelemente zusammenfaßt, die einen
- 15 besonderen Bezug zueinander haben (s.o.);
- ein betrachtetes N-Gramm (h, w) endet auf ein bestimmtes Bigramm (v, w) oder ein Lücken-Bigramm $(u, *, w)$ oder ein bestimmtes Trigramm (u, v, w) , etc.;
- ein betrachtetes N-Gramm (h, w) endet auf ein Bigramm (v, w) oder ein Lücken-Bigramm $(u, *, w)$, etc., wobei die Vokabularelemente u, v und w in bestimmten
- 20 vorgegebenen Wortklassen C, D und E liegen.

Neben der Ableitung aller Randwerte m_α nach Gleichung (3) aus einem vorgegebenen a priori- Sprachmodell mit Wahrscheinlichkeitswerten $p_{\text{ind}}(w | h)$ können für bestimmte Gruppen von Bedingungen α getrennt jeweils ein eigenes a priori-Sprachmodell mit Wahrscheinlichkeitswerten $p_{\text{ind}}(w | h)$ vorgegeben werden, wobei die Randwerte nach Gleichung

25 (3) in diesem Fall dann getrennt für jede Gruppe aus dem zugehörigen a priori-Sprachmodell berechnet werden. Beispiele für mögliche Gruppen können insbesondere gebildet werden durch:

- Wort-Unigramme, Wort-Bigramme, Wort-Trigramme;
- 30 - Wort-Lücken-1-Bigramme (mit einer einem einzigen Wort entsprechenden Lücke);
- Wort-Lücken-2-Bigramme (mit einer zwei Worten entsprechenden Lücke);

- Klassen-Unigramme, Klassen-Bigramme, Klassen-Trigramme;
- Klassen-Lücken-1-Bigramme;
- Klassen-Lücken-2-Bigramme.

- 5 Die Bestimmung der Sprachmodell-Parameter λ_α erfolgt hier mit Hilfe des GIS-Algorithmus, dessen Grundstruktur beispielsweise von J.N.Darroch, D. Ratcliff beschrieben wurde. Hier wird dabei ein Wert M mit

$$M = \max_{(h,w)} \left\{ \sum_{\alpha} f_{\alpha}(h,w) \right\} \quad (4)$$

10

angesetzt. Weiterhin soll N die Größe des verwendeten Trainingskorpus, d.h. die Anzahl im Trainingskorpus enthaltenen Vokabularelemente bezeichnen. Damit läßt sich der verwendete GIS-Algorithmus wie folgt schreiben:

- 15 Schritt 1: Starten mit irgendeinem Startwert $p_{\lambda}^{(0)}(w|h)$

Schritt 2: Aktualisierung der Randwerte im n-ten Durchlauf der Iterationsschleife:

$$m_{\alpha}^{(n)} = \sum_{(h,w)} p_{\lambda}^{(n)}(w|h) \cdot N(h) \cdot f_{\alpha}(h,w) \quad (5)$$

20

wobei $p_{\lambda}^{(n)}(w|h)$ aus den in Schritt 3 ermittelten Parametern $\lambda_{\alpha}^{(n)}$ durch Einsetzen in Formel (1) berechnet wird.

Schritt 3: Aktualisierung der Parameter λ_{α} :

25

$$\lambda_{\alpha}^{(n+1)} = \lambda_{\alpha}^{(n)} + \frac{1}{M} \cdot \log \left(\frac{m_{\alpha}}{m_{\alpha}^{(n)}} \right) - \frac{1}{M} \cdot \log \left(\frac{M \cdot N - \sum_{\beta} m_{\beta}}{M \cdot N - \sum_{\beta} m_{\beta}^{(n)}} \right) \quad (6)$$

wobei der zuletzt subtrahierte Term für den Fall entfällt, daß für M gilt:

$$M = \sum_{\beta} f_{\beta}(h, w) \quad \forall (h, w) \quad (7)$$

m_{α} bzw. m_{β} (β ist nur eine andere Laufvariable) sind die gemäß Formel (3) auf der Basis der Wahrscheinlichkeitswerte $p_{ind}(w | h)$ geschätzten Randwerte.

5

Schritt 4: Fortsetzung des Algorithmus mit Schritt 2 bis zur Konvergenz des Algorithmus.

10

Dabei wird unter Konvergenz des Algorithmus verstanden, dass der Betrag der Differenz zwischen dem geschätzten m_{α} aus Formel (3) und dem iterierten Wert $m_{\alpha}^{(n)}$ kleiner als ein vorgegebener und hinreichend kleiner Grenzwert ε ist.

15

Alternativ zur Verwendung des GIS-Algorithmus kann jedes Verfahren eingesetzt werden, das die Maximum-Entropie-Lösung zu vorgegebenen Randbedingungen berechnet, beispielsweise das Improved Iterative Scaling Verfahren, welchen von S.A. Della Pietra, V. J. Della Pietra, J. Lafferty (s. o. am angeführten Ort) beschrieben wurde.

PATENTANSPRÜCHE

- 5 1. Verfahren zur Erzeugung eines Maximum-Entropie-Sprachmodells für ein Sprach-
erkennungssystem, bei dem
durch Auswertung eines Trainingskorpus erste Wahrscheinlichkeitswerte $p_{ind}(w|h)$ für
N-Gramme mit $N \geq 0$ gebildet werden;
eine Schätzung von zweiten Wahrscheinlichkeitswerten $p_\lambda(w|h)$, die Sprachmodell-
10 werte des Maximum-Entropie-Sprachmodells repräsentieren, in Abhängigkeit von den
ersten Wahrscheinlichkeitswerten durchgeführt wird;
Randwerte m_α entsprechend der Gleichung

$$m_\alpha = \sum_{(h,w)} p_{ind}(w|h) \cdot N(h) \cdot f_\alpha(h,w)$$

- bestimmt werden, wobei $N(h)$ die Auftrittshäufigkeit der jeweiligen Vergangenheit h
15 im Trainingskorpus und $f_\alpha(h,w)$ eine Filterfunktion ist, die nur für bestimmte durch
den Index α gekennzeichnete und a priori vorgegebene N-Gramme einen von Null
verschiedenen Wert und ansonsten den Wert Null aufweist;
eine Iteration von Sprachmodellwerten des Maximum-Entropie-Sprachmodells solange
durchgeführt wird, bis im n-ten Iterationsschritt nach der Formel

$$20 \quad m_\alpha^{(n)} = \sum_{(h,w)} p_\lambda^{(n)}(w|h) \cdot N(h) \cdot f_\alpha(h,w)$$

ermittelte Werte $m_\alpha^{(n)}$ nach einem vorgebbaren Konvergenzkriterium hinreichend
genau den Randwerten m_α angenähert sind.

2. Verfahren nach Anspruch 1,
25 dadurch gekennzeichnet,
dass zur Iteration der Sprachmodellwerte des Maximum-Entropie-Sprachmodells der GIS-
Algorithmus verwendet wird.

3. Verfahren nach Anspruch 1 oder 2,

dadurch gekennzeichnet,

dass ein backing-off-Sprachmodell zur Lieferung der ersten Wahrscheinlichkeitswerte vorgesehen ist.

5

4. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

dass zur Berechnung der Randwerte m_α für verschiedene Untergruppen, die Gruppen bestimmter α zusammenfassen, verschiedene erste Wahrscheinlichkeitswerte $p_{\text{ind}}(w | h)$

10

verwendet werden.

5. Spracherkennungssystem mit einem nach einem der Ansprüche 1 bis 4 erzeugten Sprachmodell.

ZUSAMMENFASSUNG

Verfahren zur Erzeugung eines Maximum-Entropie-Sprachmodells

Die Erfindung betrifft ein Verfahren zur Erzeugung eines Maximum-Entropie-Sprachmodells für ein Spracherkennungssystem.

5

Um eine Verbesserung der statistischen Eigenschaften des erzeugten Sprachmodells zu erreichen, wird vorgeschlagen, dass

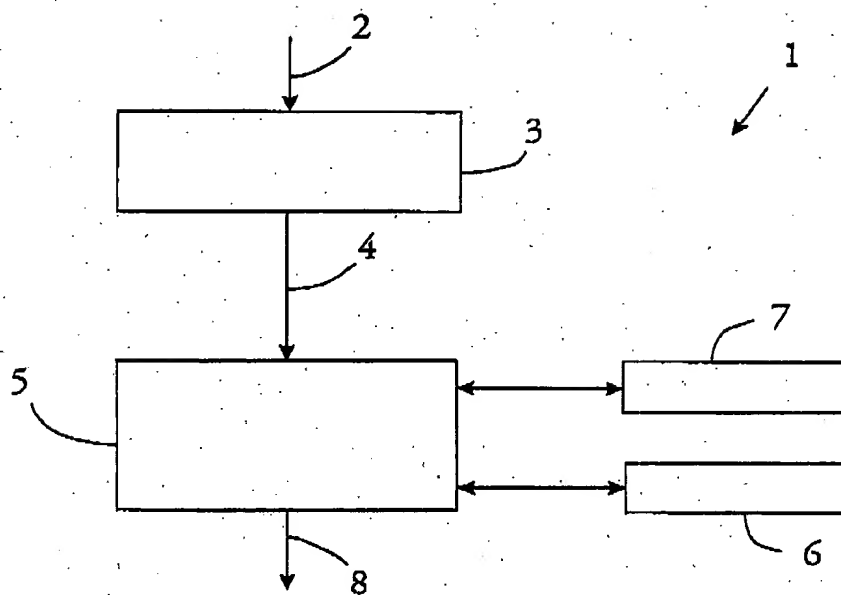
- durch Auswertung eines Trainingskorpus erste Wahrscheinlichkeitswerte $p_{md}(w|h)$ für N-Gramme mit $N \geq 0$ gebildet werden;
- 10 - eine Schätzung von zweiten Wahrscheinlichkeitswerten $p_\lambda(w|h)$, die Sprachmodellwerte des Maximum-Entropie-Sprachmodells repräsentieren, in Abhängigkeit von den ersten Wahrscheinlichkeitswerten durchgeführt wird;
- Randwerte m_α entsprechend der Gleichung

$$m_\alpha = \sum_{(h,w)} p_{md}(w|h) \cdot N(h) \cdot f_\alpha(h,w)$$

- 15 - bestimmt werden, wobei $N(h)$ die Auftretshäufigkeit der jeweiligen Vergangenheit h im Trainingskorpus und $f_\alpha(h,w)$ eine Filterfunktion ist, die nur für bestimmte durch den Index α gekennzeichnete und a priori vorgegebene N-Gramme einen von Null verschiedenen Wert und ansonsten den Wert Null aufweist;
- eine Iteration von Sprachmodellwerten des Maximum-Entropie-Sprachmodells solange
- 20 durchgeführt wird, bis im n -ten Iterationsschritt nach der Formel

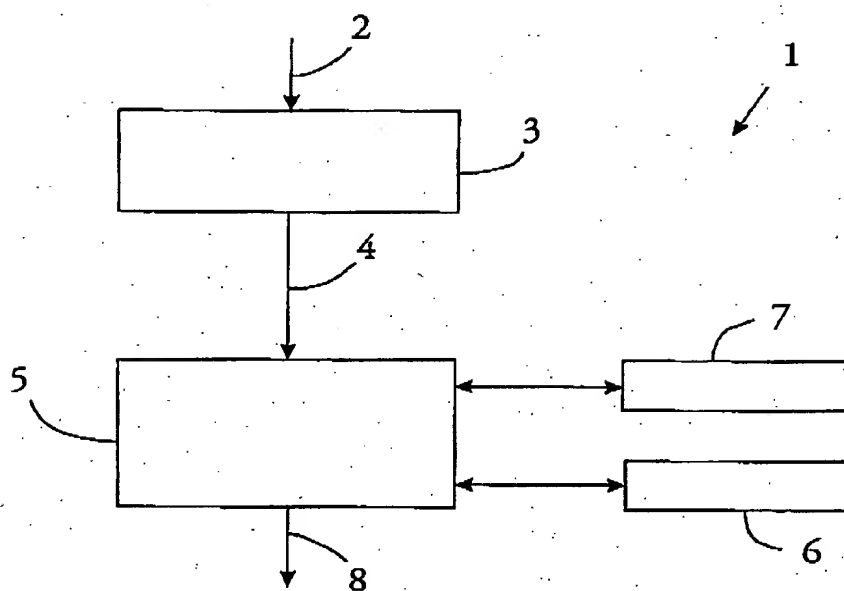
$$m_\alpha^{(n)} = \sum_{(h,w)} p_\lambda^{(n)}(w|h) \cdot N(h) \cdot f_\alpha(h,w)$$

ermittelte Werte $m_\alpha^{(n)}$ nach einem vorgebbaren Konvergenzkriterium hinreichend genau den Randwerten m_α angenähert sind.



PHD99-177

1/1



PHD99-177